A COMPUTATIONAL APPROACH TO FACTORING LARGE INTEGERS

NELSON PETULANTE

Department of Mathematics
Bowie State University
Bowie, MD 20715
USA

Abstract

To factor an integer N, given that it is equal to the product of two primes p and q, it suffices to find an integer $d < \frac{1}{2}N$ satisfying the test "$C(2\sqrt{Nd})^2 - 4Nd$ is a perfect square", where C denotes the integer ceiling function. In this approach, the factorization problem equates to the problem of designing an optimal data base $\mathcal{D}$ of values d to be tested.

KEY WORDS AND PHRASES. Divisors, factors, factoring, factorization, prime factors, large integers.

1991 AMS SUBJECT CLASSIFICATION CODES. 11D09, 14G05.

1. THE BASIC TEST

If x is a positive real number, let $C(x)$ denote the smallest integer greater than or equal to x (the integer ceiling function). Let the function $f(x)$ be defined by the formula $f(x) = C(2\sqrt{x})^2 - 4x$ and let $f^2(x)$ denote the composition $f(f(x))$.

PROPOSITION 1.1. Let n be a positive integer. Then $f(n) = 0$ if and only if n is a perfect square ($\sqrt{n}$ is a whole number).

PROOF. If $f(n) = 0$, then $C(2\sqrt{n})^2 = 4n$. Thus, $C(2\sqrt{n}) = 2\sqrt{n}$ is an *even* integer. Therefore, $\sqrt{n}$ is an integer. Conversely, if $\sqrt{n}$ is an integer, then so is $2\sqrt{n}$. Thus, $C(2\sqrt{n}) = 2\sqrt{n}$, which means that $C(2\sqrt{n})^2 = 4n$ and $f(n) = 0$.

PROPOSITION 1.2. Let n be a positive integer. Then $f^2(n) = 0$ if and only if there exist positive integers u and v such that $n = uv$ and $|\sqrt{u} - \sqrt{v}| \leq 1$.

PROOF. Suppose that $f^2(n) = 0$. By Proposition 1.1, $f(n) = t^2$ for some positive integer t. Thus, $4n = C(2\sqrt{n})^2 - t^2$. Note that $C(2\sqrt{n})$ and t must be both even or both odd. Therefore, if we let $u = \frac{1}{2}(C(2\sqrt{n}) + t)$ and $v = \frac{1}{2}(C(2\sqrt{n}) - t)$, then both u and v are integers and $n = uv$. To show that $|\sqrt{u} - \sqrt{v}| \leq 1$, observe that $u + v = C(2\sqrt{n})$, so that $u + v - 2\sqrt{uv} = C(2\sqrt{n}) - 2\sqrt{n} \leq 1$. That is,

$|\sqrt{u}-\sqrt{v}|^2 \le 1$, which implies $|\sqrt{u}-\sqrt{v}| \le 1$. Conversely, suppose $n = uv$ where $|\sqrt{u}-\sqrt{v}| \le 1$. By Proposition 1.1, to show that $f^2(n) = 0$, it suffices to show that $f(n) = t^2$ for some integer t. If $u = v$, then n is a perfect square and there is nothing to prove. So we may assume that $u \ne v$. Let $t = |u - v|$. Then $t^2 = (u+v)^2 - 4uv$. Since $|\sqrt{u}-\sqrt{v}| \le 1$, it follows that $u - 2\sqrt{uv} + v \le 1$. Therefore, since $u \ne v$, we have $2\sqrt{uv} \le u + v \le 2\sqrt{uv} + 1$, so that $u + v = C(2\sqrt{n})$ and $t^2 = C(2\sqrt{n})^2 - 4n = f(n)$. Thus, $f^2(n) = 0$.

COROLLARY 1.3. Suppose N is the product of two primes p and q where $|\sqrt{p}-\sqrt{q}| \le 1$. Then the prime factors can be recovered explicitly in terms of N by way of the formulas: $p = \frac{1}{2}(C(2\sqrt{N}) + t)$ and $q = \frac{1}{2}(C(2\sqrt{N}) - t)$, where $t = \sqrt{C(2\sqrt{N})^2 - 4N}$.

EXAMPLE 1.4. Take $N = 176039$. Then $t = \sqrt{C(2\sqrt{N})^2 - 4N} = 38$, $p = \frac{1}{2}(C(2\sqrt{N}) + t) = 439$ and $q = \frac{1}{2}(C(2\sqrt{N}) - t) = 401$. Thus, the factorization $N = (439)(401)$ follows instantly from the fact that $\sqrt{439} - \sqrt{401} \le 1$.

PROPOSITION 1.5. To factor an integer N, given that it is equal to the product of two primes p and q, it suffices to find an integer $d < \frac{1}{2}N$ satisfying the test $f^2(Nd) = 0$. Then $Nd = uv$, where $u = \frac{1}{2}(C(2\sqrt{Nd}) + t)$, $v = \frac{1}{2}(C(2\sqrt{Nd}) - t)$ and $t = \sqrt{C(2\sqrt{Nd})^2 - 4Nd}$. The prime factors p and q can be recovered separately as factors of u and v through the formulas $p = \gcd(N, u)$ and $q = \gcd(N, v)$.

PROOF. As in Proposition 1.2, with $n = Nd$, we have $f^2(Nd) = 0$. Thus, Nd factors as $Nd = uv$ where $u = \frac{1}{2}(C(2\sqrt{Nd}) + t)$, $v = \frac{1}{2}(C(2\sqrt{Nd}) - t)$ and $t = \sqrt{C(2\sqrt{Nd})^2 - 4Nd}$. It remains to show that p and q are factors of u and v separately. A rough estimate is enough to prove this. Since, by Proposition 1.2, $\sqrt{u} - \sqrt{v} = \delta \le 1$, we see $v \ge u - 2\sqrt{u}$, so that, at least whenever $u \ge 16$, we have $Nd = uv \ge u^2 - 2u^{\frac{3}{2}} \ge \frac{1}{2}u^2$. If p and q both were factors of u, we would have $u = apq = aN$ for some integer a which would imply that $Nd \ge \frac{1}{2}a^2N^2$ or $d \ge \frac{1}{2}a^2N$, contradicting the assumption that $d < \frac{1}{2}N$. A similar contradiction occurs if we suppose that p and q both divide v.

EXAMPLE 1.6. Take $N = 1110757$ and $d = 170$. Then $f^2(Nd) = 0$. Employing the formulas in Proposition 1.5, we get $t = 23$, $u = 13753$ and $v = 13730$. Thus $N = pq$, where $p = \gcd(N, u) = 1373$, $q = \gcd(N, v) = 809$.

PROPOSITION 1.7. Suppose $N = pq$ where p and q are distinct primes. An integer $d < \frac{1}{2}N$ satisfies the test $f^2(Nd) = 0$ if and only if d factors as $d = xy$ where $|\sqrt{py} - \sqrt{qx}| \le 1$.

PROOF. Suppose d satisfies the test $f^2(Nd) = 0$. Then, as in Proposition 1.2, $Nd = uv$ where $|\sqrt{u}-\sqrt{v}| \le 1$. By Proposition 1.5, we may assume that u is a multiple of p, say $u = py$, and v is a multiple of q, say $v = qx$. Thus, $d = xy$ and $|\sqrt{py} - \sqrt{qx}| \le 1$. Conversely, suppose $d = xy$ is found to satisfy the inequality $|\sqrt{py} - \sqrt{qx}| \le 1$. Then, by Proposition 1.2, $Nd = pyqx$ satisfies the test $f^2(Nd) = 0$.

2. A FACTORIZATION STRATEGY

Given $N = pq$ where the distinct prime factors p and q are unknown, our objective is to find an

integer d to satisfy the test $f^2(Nd) = 0$. By Proposition 1.7, the test $f^2(Nd) = 0$ is successful if any one of the factorizations of d as $d = xy$ satisfies the inequality $|\sqrt{py} - \sqrt{qx}| \leq 1$. Thus, for a test value d, the likelihood of success increases as $\tau(d)$, the number of divisors of d, increases. To formalize this observation, we introduce a new function defined on finite sets of integers called the "yield function".

DEFINITION 2.1. Let d be a positive integer. The *yield* of d, denoted $Y(d)$, is the number of distinct fractions $0 < \frac{x}{y} < 1$ in lowest terms such that $xyz^2 = d$ for some integer z. If $S = \{d_1, d_2, \ldots, d_k\}$ is a set of test values, then the yield of S, denoted $Y(S)$, is the number of distinct fractions $0 < \frac{x}{y} < 1$ in lowest terms such that $xyz^2 \in S$ for some integer z.

EXAMPLE 2.2. Let $d = 12$. The set of distinct fractions $0 < \frac{x}{y} < 1$ such that $xyz^2 = 12$ is $\{\frac{1}{12}, \frac{1}{3}, \frac{3}{4}\}$. Note that the fraction $\frac{1}{3}$ corresponds to the factorization $12 = (2)(6) = (1)(3)(2^2)$. Thus $Y(12) = 3$.

EXAMPLE 2.3. Let $S = \{5, 12, 20\}$. The set of distinct fractions $0 < \frac{x}{y} < 1$ such that $xyz^2 \in S$ is $\{\frac{1}{20}, \frac{1}{12}, \frac{1}{5}, \frac{1}{3}, \frac{3}{4}, \frac{4}{5}\}$. Thus $Y(S) = 6$. Note that the factorization $5 = (1)(5)$ contributes nothing to the yield of S in view of the factorization $20 = (2)(10)$. In fact, the yield of S is the same as the yield of the subset $S' = \{12, 20\}$.

Now, let $\mathcal{D}$ denote a finite data base of test values, say $\mathcal{D} = [d_1, d_2, \ldots, d_m]$, structured as a list of integers in ascending order. By definition, the *cost* of factoring the integer N relative to $\mathcal{D}$ is the number of values of $d \in \mathcal{D}$ which need to be tested before a successful value (satisfying $f^2(Nd) = 0$) is found. Obviously, our main objective is to construct a data base which minimizes the cost of factoring any given N of the form $N = pq$. Intuitively, at least, it appears fairly evident that some data bases will be more effective than others. Qualitatively speaking, the cost of factoring N relative to $\mathcal{D}$ should decrease as the yield $Y(\mathcal{D})$ increases.

At this stage, we will attempt to supply at least a rough estimate of the effectiveness of a data base. Suppose $N = pq$, where $p < q$. In $\mathcal{D}$ we want to find $d = xy$ such that $\sqrt{py} - \sqrt{qx} \leq 1$. This is equivalent to

$$\sqrt{\frac{p}{q}} - \sqrt{\frac{x}{y}} \leq \frac{1}{\sqrt{qy}}. \tag{1}$$

This inequality is satisfied if the set of fractions $\frac{x}{y}$ in the interval $[0, 1]$ is so numerous that, for at least one of them, $\sqrt{\frac{x}{y}}$ comes within a distance of $\frac{1}{\sqrt{qy}}$ of the fixed quantity $\sqrt{\frac{p}{q}}$. This will have a high probability of happening if

$$Y(\mathcal{D}) \geq \sqrt{qy}. \tag{2}$$

At this point, we need to formulate a reasonable estimate for $\sqrt{qy}$. However, without any specific knowledge of the structure of $\mathcal{D}$ this is difficult to do. We turn then to a discussion of some specific

choices of data base $\mathcal{D}$.

3. SOME SPECIAL DATA BASES

The simplest data base is a list of consecutive integers starting at 1. Let $\mathcal{D}_0(m) = [1, 2, 3, \ldots, m]$. To refine the inequality (2), note that $y \le m$. However, the median divisor of a typical $d \in \mathcal{D}$ is $\sqrt{d}$, the maximum of which is $\sqrt{m}$. It follows that $\sqrt{m}$ is a good candidate to represent y in the inequality $Y(\mathcal{D}) \ge \sqrt{qy}$. Thus, for the data base $\mathcal{D}_0(m)$, (2) becomes

$$Y(\mathcal{D}_0(m)) \ge \sqrt{q}\sqrt[4]{m}. \tag{3}$$

The explicit dependence on q can be removed by setting $R = \frac{q}{p} > 1$, so that $\sqrt{q} = \sqrt[4]{R}\sqrt[4]{N}$. Then (3) becomes

$$Y(\mathcal{D}_0(m)) \ge \sqrt[4]{mNR}. \tag{4}$$

A sharp lower bound estimate of $Y(\mathcal{D}_0(m))$ is given by m itself. To see this, note that $\sum_{d \in \mathcal{D}_0(m)} Y(d)$ is bounded above by $\sum_{1 \le k \le m} \tau(k) = O(n \ln(n))$ [1]. Thus, the data base $\mathcal{D}_0(m)$ has a good chance of factoring N if $m \ge \sqrt[4]{mNR}$. Equivalently, $\mathcal{D}_0(m)$ has a good chance of factoring N if $N \le \frac{m^3}{R}$. Often, a successful value of $d \in \mathcal{D}_0(m)$ is found well before the entire data base is exhausted. This is borne out by extensive numerical experiments using MAPLE.

A more interesting type of data base consists of the set of divisors of a given integer B. Let $\mathcal{D}_1(B) = [d_1, d_2, \ldots, d_m]$, where the d_j are the divisors of B arranged in increasing order. Thus, $m = \tau(B)$ and the largest element in $\mathcal{D}_1(B)$ is $d_m = B$. Let $B = p_1^{r_1} p_2^{r_2} p_3^{r_3} \ldots p_k^{r_k}$ (prime power factorization), then $m = (r_1 + 1)(r_2 + 1) \ldots (r_k + 1)$ and $Y_1(B) = Y(\mathcal{D}_1(B)) = (2r_1 + 1)(2r_2 + 1) \ldots (2r_k + 1)$. Thus, a good likelihood exists of factoring N provided that $Y_1(B) \ge \sqrt[4]{(N)}\sqrt[4]{(R)}\sqrt[8]{(B)}$.

Some interesting choices for B (evidenced by extensive numerical experiments using MAPLE):

$B = n!$

$B = (2)(3)(5) \ldots (p_k)$ (product of first k primes).

$B = \mathrm{lcm}(1, 2, 3, \ldots, m)$ (lcm of first m integers).

Acknowledgment: The author gratefully acknowledges the input of Steve Huntsman who found several significant errors and deficiencies in the previous version of this article. Any remaining errors or deficiencies are due solely to the author's negligence.

References

[1] H. E. Rose. *A Course in Number Theory.* Oxford University Press, 1994.

[2] S. S. Wagstaff, Jr. *Cryptanalysis of Number Theoretic Ciphers.* CRC Press, 2002.